投资者情绪、股市流动性与波动性的时变研究[1]

——基于深度学习 BERT 模型和 TVP-VAR 模型

张宸瑞，吴心怡，邓海露，张荟蔚
（华南理工大学 经济与金融学院，广州 510006）

【摘要】基于 2018 年 1 月 1 日至 2019 年 12 月 31 日东方财富网深证成指股吧的评论数据，本文通过使用深度学习 BERT 模型提取了其中蕴含的投资者情绪，并应用 TVP-VAR 模型对投资情绪、股市流动性以及波动性三者之间的时变联动关系进行了研究。实验结果表明,投资者情绪对股市流动性和波动性的冲击更为强烈，而反向的影响虽然相对较小，但其随股市状态变化更为显著。此外，在所有情况下，短期的响应都比中长期更显著，且影响具有非对称性，市场下行时期的冲击更为强烈。

【关键词】 BERT 模型；TVP-VAR 模型；深证成指；投资者情绪；股市流动性；股市波动性

A time-varying study of investor sentiment, stock market liquidity and volatility

——Based on deep learning BERT model and TVP-VAR model

ZHANG Chenrui, WU Xinyi, DENG Hailu, ZHANG Huiwei
(School of Economics and Finance, South China University of Technology, Guangzhou 510006, China)

[Abstract] Based on the review data of Shenzhen stock index BBS from January 1, 2018 to December 31, 2019, this paper extracts the investor sentiment contained in it by using the deep learning BERT model, and studies the time-varying linkage relationship among investment sentiment, stock market liquidity and volatility by using TVP-VAR model. The experimental results show that investor sentiment has a stronger impact on the liquidity and volatility of the stock market, while the reverse impact is relatively small, but it changes more significantly with the state of the stock market. In addition, in all cases, the short-term response is more significant than that in the medium and long term, and the impact is asymmetric, and the impact in the market downturn is stronger.

[Key words] Bert model; TVP-VAR model; Shenzhen Stock Index; Investor sentiment; Stock market liquidity; Stock market volatility

[1] 本文系国家大学生创新创业训练计划项目资助（项目编号：202110561076）的研究成果之一

一、 引言

股市的相对稳定是股市发展壮大的基础，它有利于国内外投资者的进入，也有利于避免金融危机的爆发。股市的异常波动将会使大部分投资者对市场失去信心，也会使作为资产配置指标的股价信号失真，进而增加整个金融体系的系统性风险。相对于欧美国家成熟的股市，中国股市波动性较大。2014 年至 2015 年上证综指暴涨暴跌事件引发了股票市场和投资者的反思，其成因可以从行为金融和分形市场理论两个角度来解释，前者突出投资者的行为和情绪的重要影响，后者强调市场流动性因素。由此可见，投资者情绪和市场流动性的动态变化已经成为市场波动性的重要影响因素。随着股票交易的互联网化和大众化，对投资者情绪、股票市场流动性及波动性间互动机制进行探究的重要性日益凸显。

本文预期的边际贡献为：（1）对当前计算机文本处理领域最前沿的深度学习模型之一——BERT 预训练模型进行微调，实现一个更适合更可靠判断中文投资者文本情绪的 BERT 模型，并将其开放开源以供后续研究者共同使用和改进；（2）当前学界中较多文献探究投资者情绪与股票市场流动性、股票市场波动性其中一者的影响关系，但较少文献研究三者之间的联动关系，本文通过使用 TVP-VAR 模型对其进行研究，以期进一步丰富学界相关领域的研究。

二、 文献综述

近年来，随着股票市场交易的互联网化和大众化，探究投资者情绪对市场的影响机制在行为金融研究领域的重要性日益凸显。通过阅读文献综述，本文将项目的研究工作分成了探究“投资者情绪测度”以及“投资者情绪与股票市场的联动关系”的两部分，并基于此进一步细化了文献阅读和研究学习的方向。

（一） 投资者情绪

直接调查法需要耗费研究者大量的时间精力，且样本容量有限偏误较大，在近期的研究中已较少被使用。市场代理指标法常通过选择封闭基金折价、换手率等代理变量并使用主成分分析法构建投资者情绪代理指标。该方法较易实现且研究效果较为显著，因此仍被广泛使用。但它也存在只能间接反映投资者情绪的缺陷。而词典法则是通过将金融文本中出现的情绪词与词典进行对照进而得到加权的投资者情绪指标。该方法能直接反映投资者情绪，极大地促进了国内外对金融文本分析的研究，但也存在情绪指标受词典质量的影响较大的问题，且构建词典的工作量比较繁琐。随着计算机算力的提升和自然语言处理领域的算法不断更新，许多学者开始借用支持向量机（SVM）、递归神经网络（RNN）和长短期记忆单元（LSTM）等模型算法并利用机器学习的方法直接提取金融文本中的投资者情绪，并构建指标，此类研究随着机器学习模型算法的改进不断提高准确率。Huicheng Liu 利用双向 LSTM 对新闻文本进行编码并搜捕上下文情绪信号特征；Yao Qin 利用非线性自回归模型，构建出一种基于双阶段递归神经网络时间序列的预测方法来研究投资者情绪。目前，机器学习法是投资者情绪测度领域最前沿的方向，因此本文考虑采用机器学习的方法研究投资情绪。

情感分类是自然语言处理任务中的一个子任务，其本质上是对带有情感色彩的主观性文本进行分析、处理、归纳和推理的过程，而本文研究的是对股吧情绪的分析处理，情感分类有利于对股市舆情监测进行合理检测。

随着计算机硬件的发展，一种网络结构更深、更复杂的深度学习模型——BERT 由此而被设计出来。研究发现 BERT 更多地用于媒体、政治评论研究，例如 Guoshuai Zhang 等（2020）利用 BERT 语言模型分析《纽约时报》发布的新闻，预测美国未来短期内的政策变化；Jiaqi Hou 等（2020）结合 transformer 的双向编码器的 BERT 模型，提出一种新的 BERT-att，并证明这种模型比领先的基线模型更有效，可用于安全风险评估等领域。然而在金融研究领域，仍鲜有学者在研究中使用 BERT 模型，也较少有学者将其调

试好后的可以直接使用的情绪判断机器模型进行开源，研究希望能为本领域填补上这方面的差距和空白，以便未来的研究能在该方向上持续推进和优化，本文的研究对当前计算机文本处理领域最前沿的深度学习模型之一——BERT 预训练模型进行微调，实现一个更适合更可靠判断中文投资者文本情绪 BERT 模型。

（二） 市场流动性

股市流动性是指股市迅速地、低成本地执行大量交易并且不会造成价格大幅变化的现象。流动性包括几个方面的内容：一是交易的即时性；二是交易的低成本性；三是可交易的股票数量巨大；四是价格偏低程度小。目前研究中用的的流动性指标众多，可以根据研究方向的需要构建不同的流动性指标。杨朝军(2008)借鉴 Amivest 流动性比率与 Hui-Heubel 流动性比率思想的基础上,提出新的流动性指标；曾志坚和罗长青(2008)以换手率为流动性变量 ， 利用一年上交所数据研究了两个市场流动性之间的联动关系。研究发现股票市场和债券市场月度流动性之间不存在领先滞后关系 ；王茵田等（2010）对股票市场和债券市场的流动性溢出效应进行研究，构建出基于价格影响的流动性衡量方法的流动性指标；韩金晓等（2017）用 Crowin 和 Schultz（2012）价差类指标可以有效地度量中国股票市场流动性。在股票市场，换手率是最简单的流动性指标。

（三） 市场波动性

股票市场波动性是指由于经济因素、政策因素和市场参与者自身因素的变化引起对市场预期的变化而产生的股票价格的变化。股票市场波动受基本面、资金面、投资者情绪等多方面因素的影响。对于金融市场来说，波动性是非常重要的特征之一，它反映金融资产回报随之间的变化。对于波动性的测算方法，已有文献中较为普遍的方法有：收益率的方差或标准差、收益率的条件方差、震荡幅度等。Engle（1982）的研究中将方差和条件方差进行区分，建立了 ARCH 模型，开创了波动性研究的新思路。Bollerslev（1896）在此基础上提出了 GARCH 模型。黄海南和钟伟（2007）提出运用 GARCH 类模型来预测上证指数的收益波动率。戴雯（2018）利用 ARMA-GARCH 模型族对 ST 类股票收益率进行实证分析。现有文献对于股票市场波动性研究方面揭示了证券价格波动的聚集性、持续性和非对称性，验证了 ARCH 族模型在证券价格波动建模中的适用性。目前国内外学者提出的多种模型从不同方面考量股市波动性，可以根据研究所需进行选择，基于实

际情况，本文参考国泰安数据库中的已实现波动率来构建股市波动性指标。

（四） 投资者情绪与股票市场流动性、股票市场波动性相关研究

在投资者情绪对流动性的影响方面，Chen 等（2009）研究了投资者情绪、市场流动性和投资者交易行为之间的联系，结果发现投资者的情绪倾向于悲观时，更多的投资者会选择卖出，减少了市场的净买入量，进而降低了市场流动性，同时个人投资者对于情绪变化的反应程度高于机构投资者。李春红和彭光裕（2011）运用格兰杰因果关系发现投资者情绪是股票流动性的格兰杰原因，投资者情绪通过影响股票流动性进而间接影响股票市场的资本化率和经济增长水平。王丹枫和梁丹（2012）发现投资者情绪变化会增加市场流动性，对股票预期收益的流动性溢价的影响也变得显著。刘晓星等（2016）发现投资者情绪会正向影响市场流动性，且投资者对于股市信息的认知能力越高时，股市流动性相对减弱。此外，研究发现投资者的积极情绪和消极情绪对于股市流动性的影响程度存在差异。杨晓兰等（2016）通过本地偏好、投资者情绪对股市的影响研究，发现本地关注对股票交易量的正向影响在积极情绪下比消极情绪下的影响程度更大。石广平等（2016）利用 TVP-SV-SVAR 模型分析投资者情绪、市场流动性对股市泡沫的影响，发现投资者情绪对股市流动性的影响，且在牛市中乐观情绪的影响比在熊市中悲观情绪的影响更加显著。上述文献从不同视角研究了投资者情绪对股市流动性的影响，并在一定程度上发现了不同类型的情绪对于股市的差异化影响以及在不同时期下的差异化表现。

在投资者情绪对股票市场波动的影响方面，Lee 等(2002)发现投资者情绪对股票市场波动性有显著影响，其中积极情绪会减弱市场波动，消极情绪会增大市场波动。张宗新和王海亮（2013）运用主成分分析法构建投资者情绪指数，分析投资者情绪与股市波动之间的关系，发现投资者情绪对市场收益率和波动率存在显著的正面影响，情绪水平越高，股市收益率变化越大，股价偏离内在价值越远，股市波动越大。胡昌生和池阳春（2013）发现理性投资者和非理性投资者都是市场波动的制造者，投资者的情绪会使市场偏离理性框架，导致异常波动。陆剑清和陈明珠（2013）提出“预期情绪溢出效应”假说，认为微观层面的投资者的心理预期激发了羊群效应，诱发了宏观层面的“股市同频共振效应”，加剧了股市波动。巴曙松（2016）指出融资融券将助长投资者情绪，并对市场波动性具有加剧效应。上述文献从不同角度出发，证实了投资者情绪是股市系统性风险因素之一，也是影响股市收益波动的重要因素。

三、 数据处理与模型构建

（一） 数据的选取

考虑研究对象的选取，需要对整个 A 股市场有一个较好的概括，能够全方位的反映投资者情绪的实际状况，本文选取东方财富网股吧作为文本数据来源，选取已实现波动率（RV_t）、换手率（$Turn_t$）作为变量来研究投资者情绪的影响。

1. 文本数据的选取

本文爬取东方财富网[2]股吧 2018 年 1 月 1 日至 2019 年 12 月 31 日的股吧数据，横向对比多个个股以及指数股吧数据，最终选择深证成指的股吧。相比上证综指以及其他指数股吧数据，深证成指股吧具有数据量合适、浏览人数多、代表性强等特征。其次，深证成指包含了在深交所上市的具有一定规模的 500 支优质公司股，涵盖了各个行业的数据。因此深证成指能很好的反应股票市场的情况。在文本处理上，遵循文本分析标准，删除了重复数据、数据中的表格、HTML 标签等非文本项目。

表 1 2018 年 1 月 1 日深证成指部分成分股

排名	代码	简称	权重（%）	总股本（万亿）
1	000333.SZ	美的集团	3.63	65.61
2	000651.SZ	格力电器	3.10	60.16
3	000858.SZ	五粮液	2.20	37.96
4	000725.SZ	京东方 A	2.19	347.98
5	000002.SZ	万科 A	1.88	110.39

注：数据来源同花顺 iFind 数据库。

表 2 每日评论统计汇总信息

时间	评论	阅读量	评论数
2019-1-2	不可否认，今日又是涨幅巨大股票甩抛机会。必须用平静心态面对涨涨跌跌	2115	0
2019-1-2	北向外资流入 A 股，净流入逾 3100 亿元，利好。	1159	2
2019-1-2	股市内外环境发生根本改变，别用陈旧熊市眼光看待股市，股市大开放股价必须步步上台阶	921	0

由于股票市值是不断变动的，因此成分股也是不断变化的，表 1 是 2018 年 1

2 根据 iResearch 2018 年的一份分析报告指出，东方财富网是中国顶级金融网站。每月有效浏览时间为 7800 万小时，占市场份额的 45%，高于前 10 家公司中其余 9 家公司的总和。

月 1 日深证成指部分成分股，但在本文研究期间内，成分股变化不大，因此可以不考虑成分股变化带来的影响。此外，我们对每日股吧爬取的信息包括时间、评论内容、阅读数和评论数等内容，如表 2 所示。

从表 3 所得到的统计信息可以看出，每天平均评论标题长度为 62.7 个字符，且内容呈右偏分布。与传统的（相对较短的）评论不同的是，股吧数据中一些字数较多的股评通常是从其他来源复制和粘贴而来的，如新闻报道和分析报告。在这里我们使用简单的处理来消除这些潜在的潜在影响离群值，仅保留少于 150 个汉字的股评。此外考虑到每日评论数参差不齐以及考虑不同阅读量对结果的影响，我们选择了按阅读量排序，由高到低每日选取了前 50 条数据进行研究。

表 3 经预处理的股吧评论文本数据的基本统计描述

	Mean	S.D	Skewness	Min	Max	Count
数据长度（字符）	62.6754	169.0293	7.9438	2	2921	39784

注：此表为每天评论长度的汇总统计信息。变量的样本均值、标准差（S.D.）、最大最小数和评论总数。该样本包含 2018 年 1 月 1 日至 2019 年 12 月 31 日样本期内深证成指股吧评论数据。

2. 股票交易数据的选取

本文选取深证成指作为研究对象，借助其每日交易行情指数进行研究，包括当天的总市值、流通市值、总股本、流通股本、换手率、市盈率、市净率额等数据。本文通过开源的 Python 数据 API——Tushare 获取，返回的结果为 Pandas.DataFrame 数据类型。

表 4 2018-2019 年各指标的基本统计描述

	Mean	S.D	Skewness	Min	Max
股市换手率（%）	1.2565	0.5268	2.3546	0.66	3.99
股市波动率（%）	1.1936	1.0730	3.6074	0.21	12.02

由表 4 可知，股市波动性与换手率平均值在 1.2%左右，且二者均呈现右偏分布，其中股市波动率较大，最大最小值之间相差约 10%可见波动较为剧烈。

（二） 模型构建

1. 基于金融语料的深度学习模型设计

深度学习时代中的 NLP（自然语言处理）预训练工作广泛使用词嵌入（Word Embedding），使用深度学习模型进行训练的时候，会将所训练的次转化为词向量作为神经网络的输入层，而在深度学习模型训练的过程中，训练结果的好坏水平很大程度取决于训练集的大小，较大的训练集可以训练出较好的词向量，目前在自然语言处理领域绝大部分的任务模型都会采用训练好的词向量。在词向量的训练过程中，词向量忽略了上下文的表意，当词汇出现一词多义的情况，往往对应的是相同的词向量。因此，2018 年 Devlin 等人（2019）提出预训练语言模型 BERT，问鼎 11 个 NLP 任务的榜单，是 NLP 领域前进的一大步。BERT 模型的结构如图 1 所示，其中 $E_1,E_2,\ldots,E_N$ 是模型的输入字符，输入字符通过双向的 Transformer 特征提取器获取文本特征，输入字符训练后输出相应的向量 $T_1,T_2,\ldots,T_N$。

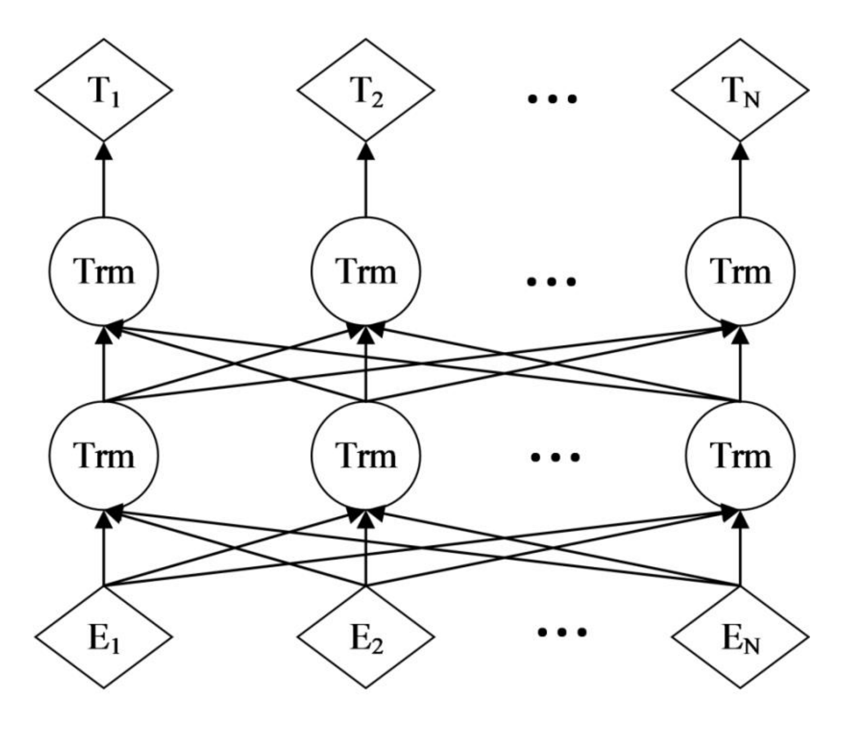


图 1 BERT 模型结构

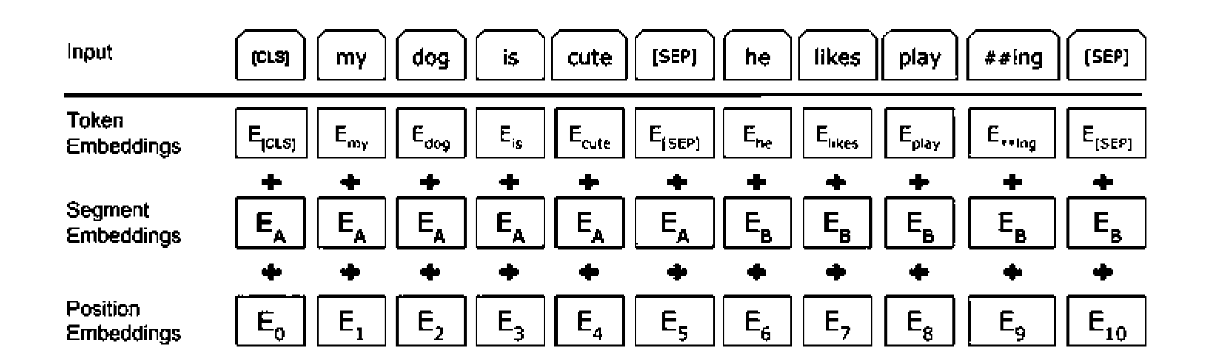


图 2 BERT 模型的训练方式

如图 2 所示，BERT 作为一个所有层都能够结合上下文语义进行训练的模型，其输入由字嵌入（Token Embeddings）、段嵌入（Segment Embeddings）以及位置嵌入（Position Embeddings）三个向量组成，与此同时，BERT 采用的是 MLM 模型

（遮盖语言模型），MLM 通过遮盖（Mask）一部分字，类似填空，然后去预测被遮盖的模型，通过迭代来达到上下文训练的目的。

BERT 本质上是两段式 NLP 模型。第一个阶段叫做：Pre-training（预训练），利用现有无标记的语料训练一个语言模型。为了提高在预测中的准确率这里采用哈工大讯飞联合实验室发布基于全词覆盖（Whole Word Masking，wwm）的中文 BERT 预训练模型 Chinese-BERT-wwm 作为预训练模型。BERT 在完成预训练后，可将其用于金融实体情感识别的任务，在情感分析中将作为下一网络的输出。根据金融文本的特殊性，在本地使用 GPU 将带情感标注的金融文本进行第二阶段 Fine-tuning（微调），就可以训练出在金融等特定领域精度更高的模型。

(1) 模型训练

本文模型选用工大讯飞联合实验室发布的 chinese_roberta_wwm_large_ext_L-24_H-1024_A-16（24-layer, 1024-hidden, 16-heads）预训练模型，采用了 24 层 Transformer，其隐层维度为 1024，多头注意力参数为 16，参数模型总大小为 330MB。模型训练方面批次大小（batch size）为 16，学习率（learning rate）为 2e-5，序列最大长度（max seq length）为 128。本文使用三分类对情绪进行标注，0 表示负向情绪、1 表示中性情绪、2 表示正向情绪，在所爬取的 2018 年 1 月 1 日至 2019 年 12 月 30 日的数据中，随机选取 4000 条数据，进行人工情绪的标注，并对所标注的数据按 8:2 的比例划分训练集及测试集，进行训练。如表 5 所示，处理的数据包括情绪（Sentiment）0 代表消极，1 代表中性，2 代表积极，Positive_Pro 指文本属于积极情绪

的概率，Negtive_Pro 表示属于消极情绪的概率。

表 5 部分样例数据

Date	Context	Sentiment	Positive_Pro	Negtive_Pro
2018-12-21	10 年一次的财富翻倍机会，你想实现财富自由吗？现在抄底让你实现梦想[大笑][牛]	2	0.99148	0.00852046
2018-12-21	上联，[牛]借东墙补西墙，年年没钱。[牛]下联，[牛]借新帐还旧账，年年亏损。	0	0.0499047	0.950095
2018-12-21	A 股离完蛋有多远？	0	0.042031	0.957969

(2) **基于股评的投资者情绪指标**

借鉴 Antweiler 和 Frank（2004）提出的基于股吧帖子分类构建看涨指标的方法，即：

$$Sent_t = \frac{M_t^{pos} - M_t^{neg}}{M_t^{pos} + M_t^{neg}} \quad (1)$$

式（1）中$M_t^c = \sum_{i\in D(t)} \omega_i \chi_i^c$ 表示类型为 c∈{pos,neu,neg}的消息在一段时间 D(t) 内的加权数量之和其中 pos 代表积极情绪，neg 代表消极情绪，neu 代表中性情绪χ_i^c是指示变量，如果消息属于某种类型 c，其值为 1，否则为 0；特别地，当权重全部等于 1 时，M_t^c 就等于 D(t) 时段内类型为 c 的消息总量，股评情绪指标$Sent_t$，介于-1 和 1 之间，表达了投资者们相对看涨的程度，该指标与帖子总数量无关。由图 3 及表 6 可以看出情绪以负向居多，负向情绪在 2018 年为集中，这与 2018 年单边下行趋势有关。

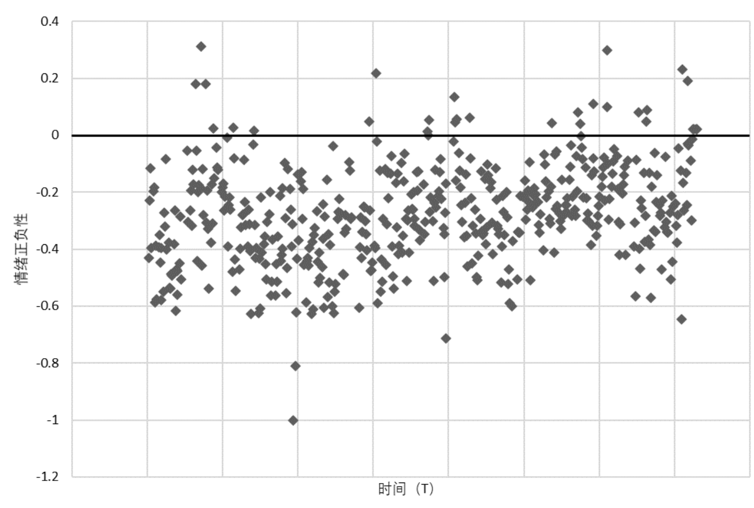


图 3 情绪分布散点图

四、 实证分析

（一） 平稳性检验

首先对投资者情绪（$Sent_t$）、已实现波动率（RV_t）以及换手率（$Turn_t$）这三项指标进行平稳性检验。实验结果表明，在 1%的显著性水平下，投资者情绪（$Sent_t$）和已实现波动率（RV_t）序列为零阶单整，而换手率（$Turn_t$）序列为一阶单整。

表 6 变量平稳性检验结果汇总表

	检验形式(C,T,L)	t 统计量	p 值
$Sent_t$	(1,0,1)	−10.89045	0.0000***
RV_t	(1,0,2)	−6.132106	0.0000***
$dTurn_t$	(1,0,4)	−14.31729	0.0000***

(C,T,L)分别表示 ADF 检验模型中的截距项、趋势项、滞后阶数；***表示在 1%的置信水平下显著

（二） TVP-VAR 模型的建立

TVP-VAR 模型假设系数和误差项的方差具有时变性，满足一阶随机游走过程。时变性使得模型能够更好地捕捉投资者情绪以及股市波动间可能存在的、非线性的、渐进性的联动变化或者是结构性的转变。然而，时变性也使得传统的似然函数法难以可靠地估计参数。Nakajima(2011)提出借助马尔可夫蒙特卡洛算法（MCMC）来克服传统估计法的缺陷。首先，根据经验设定时变参数的先验分布；其次，通过马尔可夫蒙特卡洛算法迭代抽取状态向量的随机样本。当满足相关条件时，马尔科夫链的状态转移矩阵将收敛到一个稳定的概率分布，选取收敛后、满足平稳分布的样本集即可有效地模拟时变参数的条件后验分布，进而实现参数的统计推断。

根据六项评判指标准则（LogL、LR、FPE、AIC、SC、HQ）综合的结果，判断 VAR 模型的最佳滞后阶数为 3 阶。借鉴 Nakajima(2011)和郑挺国等（2018）的研究,本文对投资者情绪（$Sent_t$）、已实现波动率（RV_t）以及换手率一阶差分序列（$dTurn_t$）建立如下时变参数向量自回归（TVP-VAR）模型。

$$\boldsymbol{A_t y_t = F_{1t} y_{t-1} + F_{2t} y_{t-2} + F_{3t} y_{t-3} + u_t, \quad u_t \sim N(0, \textstyle\sum_t \sum_t)} \tag{1}$$

其中，y_t为3×1维列向量（$Sent_t, RV_t, dTurn_t$），F_{1t}、F_{2t}、F_{3t}为3×3维时变系数矩阵，随机扰动项ε_t为3×1维列向量，能反映结构冲击，$\sum_t$为以随机扰动项标准差（$\sigma_{1t},\sigma_{2t},\sigma_{3t}$）为元素的对角矩阵。$A_t$为对角线元素为 1 的下三角矩阵，能反映通过递归识别得到的、由结构冲击带来的、不同变量间相互影响的同时关系。

对方程（1）式进行变形，并增加截距项，可以得到方程的最终形式：

$$\boldsymbol{y_t = C_t + B_{1t}y_{t-1} + B_{2t}y_{t-2} + B_{3t}y_{t-3} + e_t}, \quad \boldsymbol{e_t \sim N(0, A_t^{-1}\textstyle\sum_t \sum_t A_t'^{-1})} \qquad (2)$$

其中，$B_{it} = A_t^{-1}F_{it}$ （$i = 1,2,3$），将B_{it}的元素按行堆叠形成27×1维的列向量β_t，将A_t非对角线上的下三角元素按行堆叠形成列向量a_t。$e_t = A_t^{-1}\sum\varepsilon_t$（$\varepsilon_t \sim N(0, I_3)$），$h_t = (h_{1t}, h_{2t}, h_{3t})'$，并且满足$h_{jt} = log\sigma_{jt}^2$ （$j = 1,2,3$）。

假设方程（2）式中所有的参数满足一阶随机游走过程，并且误差项服从如下分布：

$$\boldsymbol{\beta_{t+1} = \beta_t + u_{\beta t}}, \qquad \boldsymbol{a_{t+1} = a_t + u_{at}}, \qquad \boldsymbol{h_{t+1} = h_t + u_{ht}}$$

$$\begin{pmatrix} \varepsilon_t \\ \boldsymbol{u_{\beta t}} \\ \boldsymbol{u_{at}} \\ \boldsymbol{u_{ht}} \end{pmatrix} \sim N\left(0, \begin{pmatrix} I & O & O & O \\ O & \sum_\beta & O & O \\ O & O & \sum_a & O \\ O & O & O & \sum_h \end{pmatrix}\right)$$

$$\boldsymbol{(\Sigma_\beta)_i^{-2} \sim Gamma(20, 10^{-4})}, \qquad \boldsymbol{(\Sigma_a)_i^{-2} \sim Gamma(4, 10^{-4})}, \qquad \boldsymbol{(\Sigma_h)_i^{-2} \sim Gamma(4, 10^{-4})}$$

参考郑挺国（2018）的模型设定，假设初始状态如下：

$$\boldsymbol{\beta_0 \sim N(\mu_{\beta_0}, \textstyle\sum_{\beta_0})}, \qquad \boldsymbol{a_0 \sim N(\mu_{a_0}, \textstyle\sum_{a_0})}, \qquad \boldsymbol{h_0 \sim N(\mu_{h_0}, \textstyle\sum_{h_0})}$$

$$\boldsymbol{\mu_{\beta_0} = \mu_{a_0} = \mu_{h_0} = 0}, \qquad \boldsymbol{\textstyle\sum_{\beta_0} = \sum_{a_0} = \sum_{h_0} = 10 \times I}$$

（三） 格兰杰因果检验

对投资者情绪（$Sent_t$）、已实现波动率（RV_t）以及换手率一阶差分序列（$dTurn_t$）进行格兰杰因果检验。滞后阶数与 TVP-VAR 模型一致（3 阶），实验结果如下：

表 7 格兰杰因果检验表

原假设	F 统计量	p 值
$Sent_t$不是引起$dTurn_t$变化的格兰杰原因	2.48165	0.0603
$dTurn_t$不是引起$Sent_t$变化的格兰杰原因	3.14444	0.0250
RV_t不是引起$dTurn_t$变化的格兰杰原因	9.50156	4.E-06
$dTurn_t$不是引起RV_t变化的格兰杰原因	1.16143	0.3240
RV_t不是引起$Sent_t$变化的格兰杰原因	0.87328	0.4548
$Sent_t$不是引起RV_t变化的格兰杰原因	2.84445	0.0373

检验结果表示，在 10%的显著性水平下，拒绝“投资者情绪（$Sent_t$）不是引起换手率（$dTurn_t$）变化的格兰杰原因”的原假设；在 5%的显著性水平下，拒绝“换手率（$dTurn_t$）

不是引起投资者情绪（$Sent_t$）变化的格兰杰原因”以及投资者情绪（$Sent_t$）不是引起已实现波动率（RV_t）变化的格兰杰原因”的原假设；在 1%的显著性水平下，拒绝“已实现波动率（RV_t）不是引起换手率（$dTurn_t$）变化的格兰杰原因”的原假设。即投资者情绪的前期变化能有效地解释换手率和波动率的波动；换手率的前期变化能有效地解释投资者情绪的前期变化；已实现波动率的前期变化能有效地解释换手率的波动。

（四） 实验结果分析

设定马尔科夫蒙特卡洛算法迭代次数为 11000 次，舍弃前 1000 次的“预烧”抽样，利用后 10000 次满足平稳分布的抽样进行估计。基于马尔科夫蒙特卡洛算法的 TVP-VAR 模型估计结果如下：

表 6 基于 MCMC 算法的 TVP-VAR 模型估计结果表

参数	均值	标准差	95%置信区间	Geweke 诊断值	无效影响因子
$(\Sigma_\beta)_1$	0.0023	0.0003	[0.0018,0.0029]	0.864	28.76
$(\Sigma_\beta)_2$	0.0023	0.0003	[0.0018,0.0029]	0.928	21.76
$(\Sigma_a)_1$	0.0068	0.0021	[0.0038,0.0113]	0.900	99.99
$(\Sigma_h)_1$	0.0064	0.0028	[0.0033,0.0135]	0.204	168.34
$(\Sigma_h)_2$	0.7354	0.0841	[0.5844,0.9123]	0.374	45.01

实验结果表明，所有估计的参数值都落在其 95%置信区间内，所有参数的 Geweke 诊断值都未超过 5%显著性水平的临界值，即所有参数估计的马尔科夫链都是满足收敛性的。综上可知，参数估计具备有效性特征。

1. 不同滞后期的脉冲响应分析

参考杨文祺等（2020）和任永平等（2020）的研究，本文分别选取 1 天、7 天（一个星期）以及 14 天（两个星期）的滞后期为代表来研究单位冲击变量对被冲击变量在短期、中期和长期内形成的影响。

(1) 不同滞后期投资者情绪与股市流动性间的影响机制

在整个研究时间段内，投资者情绪对股市流动性指标的脉冲响应恒为正值，并存在一定的时变效应。在 2018 年 6 月（t=100 附近）至 2019 年 7 月（t=360 附近）期间，

滞后 1 期的来自投资者情绪变动的冲击对股市流动性的影响缓慢增加。即当投资者面对股票市场的变化并经过一段时间确认进入熊市（t=20 附近）后，由于对后市走势以及持续时间的不确定性以及风险厌恶的偏好，可能会导致其频繁交易从而使得市场过度反应。这样的影响一直持续至投资者经过一段时间确认进入牛市（t=246 附近）。

此外，滞后 1 期（短期）的影响显著强于滞后 7 期和 14 期（中长期）的影响。这也反映了股票市场的信息效率，与杨文祺（2020）的研究结果一致。反向地，虽然市场流动性指标对投资者情绪的影响相对较小，但是其时变效应更显著（受股市状态变化影响更强烈）。实验结果如图 4 所示。

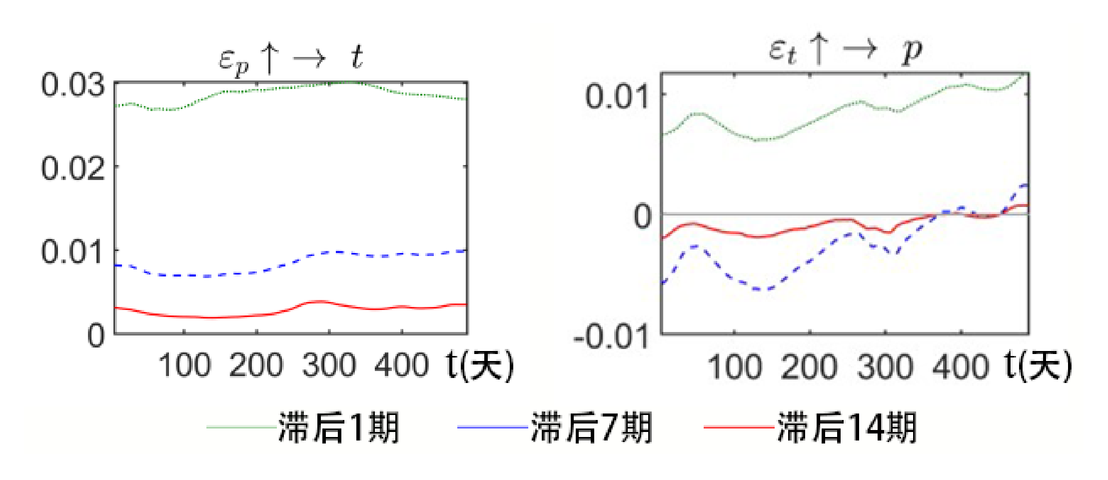


图 4 不同滞后期投资者情绪与股市流动性间的脉冲响应图

(2) 不同滞后期投资者情绪与股市波动性间的影响机制

实验结果如图 5 所示，在研究时间段内，滞后 1 期的投资者情绪对股市波动率的脉冲响应恒为负值，且熊市时冲击更强烈，中长期的影响微弱。

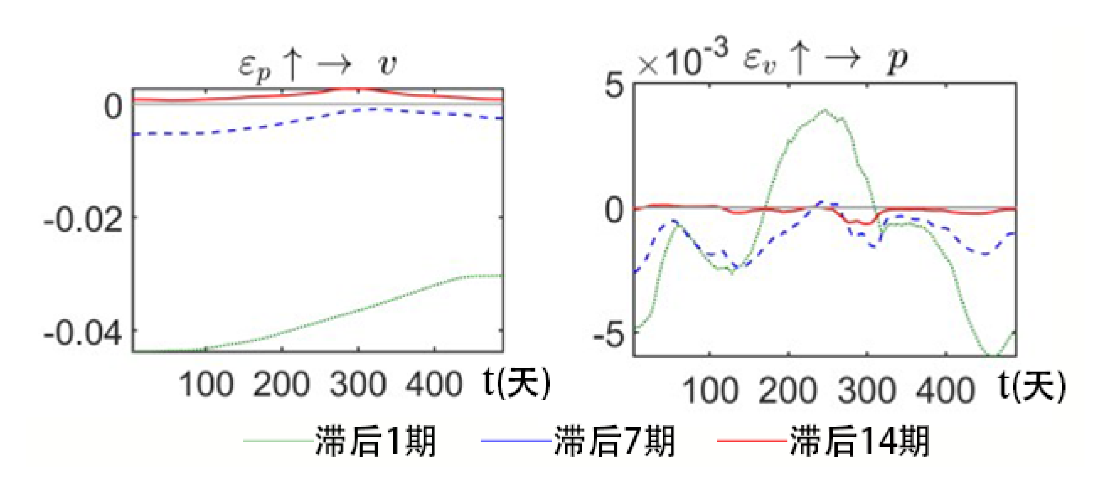


图 5 不同滞后期投资者情绪与股市波动性间的脉冲响应图

2. 不同时间点的脉冲响应分析

本文研究的时间段内覆盖了一段慢牛市的末期（2018 年 1 月 2 日至 2018 年 1 月 29 日）、一段熊市（2018 年 1 月 30 日至 2019 年 1 月 4 日）和一段新牛市的前中期（2018 年 1 月 5 日至 2019 年 12 月 31 日）。通过选取 2018 年 1 月 15 日（t=10）、2018 年 7 月 2 日（t=120）以及 2019 年 6 月 27 日（t=360）三个时间点可以具有代表性地探究在牛市末期、熊市中以及牛市前中期单位冲击变量对被冲击变量影响的差异。

(1) 不同时间点投资者情绪与股市流动性间的影响机制

在三个时间点上，投资者情绪对股市流动性指标的冲击响应恒为正值，且短期影响强于长期影响，滞后 3 期的情绪对股市流动性的影响显著减弱。牛市时投资者情绪对股市流动性的冲击稍强。实验结果支持情绪对股票市场影响具有不对称性的认识。乐观情绪更容易推动股票价格向上偏离基本价值，进而容易导致频繁交易，股价泡沫产生。反向地，市场流动性指标对投资者情绪的冲击响应在滞后 2 期内为正值，随后下降为负值，影响缓慢减弱为 0。可能的传导机制为：在投资者观察到市场交易频繁，股票换手率提高的初期，其情绪会容易随之高涨并也跟随交易，即羊群效应。个体的非理性决策容易演变成趋同的市场群体情绪和行动，从而造成频繁交易，资产价格的偏离以及市场的波动等现象。然而在投资者处理信息做出理性判断后，股市流动性变化的冲击即会减弱。此外，值得注意的是，熊市时市场流动性变化对投资者情绪的冲击显著强烈于牛市。实验结果如图 6 所示。

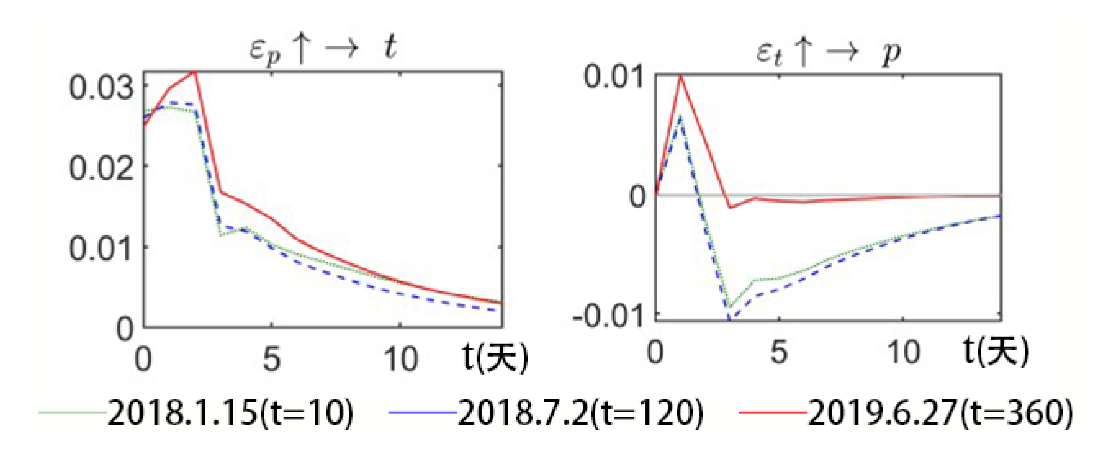


图 6 不同时间点投资者情绪与股市流动性间的脉冲响应图

(2) 不同时间点投资者情绪与股市波动性间的影响机制

实验结果表明投资者情绪在熊市时比在牛市时对股市波动性指标的影响更为强烈，且影响力大小在后 5 日内经历反复变化后逐渐衰弱。这可能与投资者接受并处理信息的能力以及信息传播的效率有关。

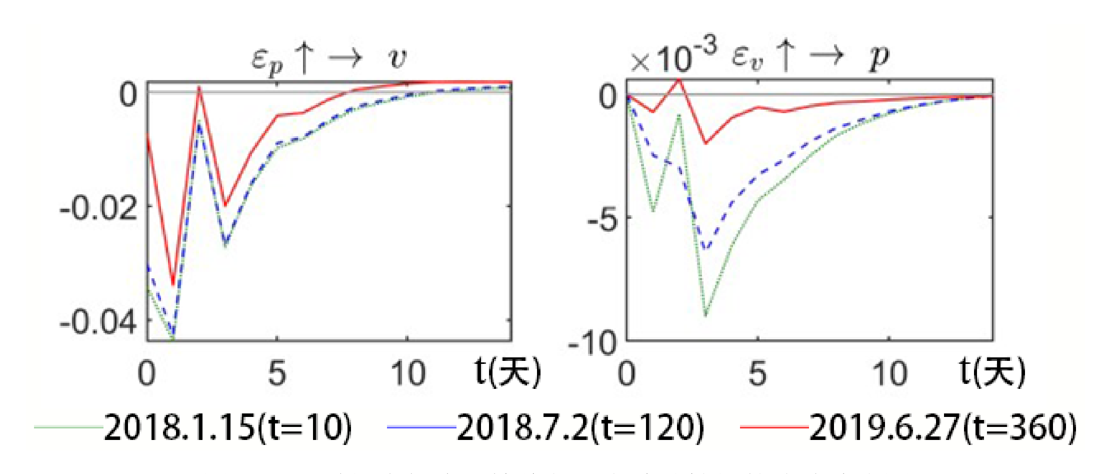


图 7 不同时间点投资者情绪与股市波动性间的脉冲响应图

五、 结论与建议

本文就投资者情绪对股市的影响做出研究，提出使用深度学习的方法研究大量的市场情绪，基于东方财富网和深证成指的数据，使用 BERT 模型构建投资者情绪，同时引入市场流动性、市场波动性指标。实验结果显示相关性较高，说明投资者情绪确实对股票市场有较大影响。剩余的结果偏离可能是由于实证分析中数据爬取不完善、股吧数据噪声较大、分析模型有微小偏差引起的。但在大的方面符合假设，实验结果也说明了该模型具有较强的稳健性。

本文的研究结论包括：1）BERT 模型能有效地挖掘出股吧中的投资者情绪。 2）TVP-VAR 模型能够有效搭建投资者情绪、流动性和波动性三者之间的关系。

基于本次研究的政策建议有：1） 中国股票市场投资者情绪对股市影响较大，尤其是个人投资者情绪容易受到股吧等社交平台情绪的影响，这对进一步完善国家市场经济制度做出要求。随着股票市场的完善，机构投资者的增多，股价对投资者情绪的敏感度会降低，市场风险也将降低。2）把发展机构投资者作为我国证券市场重要发展战略的同时,如何提高机构投资者的素质、增强其投资行为的理性化程度也是一个同等重要的问题。这一问题解决得好,才能提高个体投资者素质 ,进而提高市场效率。

本文存在着一些未完善之处有待学者进一步研究。例如：未来的工作可以尝试研究国家政策、市场、突发事件等不确定性因素影响，也可以在模型中加入更多解释变量减少噪声，尝试使用长周期样本以及更小的样本来进一步试验完善模型和参数，提高模型的性能。

参考文献:

[1] Leveraging Financial News for Stock Trend Prediction with Attention-Based Recurrent Neural Network; Huicheng Liu
[2] A Dual-Stage Attention-Based Recurrent Neural Network for Time Series Prediction; Yao Qin et al
[3] Learning to Predict U.S. Policy Change Using New York Times Corpus with Pre-Trained Language Model; Guoshuai Zhang
[4] A Neural Relation Extraction Model for Distant Supervision in Counter-Terrorism Scenario; Jiaqi Hou
[5] 廖士光,杨朝军.中国股票市场流动性价值研究——双重上市 A、B 股经验证据[J].南方经济,2008(12):64-78.
[6] 曾志坚,罗长青.股票与债券市场流动性联动的实证研究[J].财经理论与实践,2008(04):45-49.
[7] 王茵田,文志瑛.股票市场和债券市场的流动性溢出效应研究[J].金融研究,2010(03):155-166.
[8] 韩金晓,吴卫星.流动性与资产定价——基于一个新的价差类流动性度量指标的实证研究[J].上海金融,2017(12):51-57.DOI:10.13910/j.cnki.shjr.2017.12.008.
[9] Engle,R..Autoregressive Conditional Heteroskedasticity with estimates of the variance of United Kingdom inflation[J].Econometrica, 1982,(50):987-1007.
[10] Bollerslev T.Generalied autoregressive conditional heteroskedasticity [J].Journal of Econometrics,1986,(31):307-327.
[11] 戴雯.ST 类股票收益率波动性的实证分析——基于 ARMA-GARCH 模型族[J].当代会计,2018(10):9-10.
[12] 黄海南,钟伟.GARCH 类模型波动率预测评价[J].中国管理科学,2007(06):13-19.DOI:10.16381/j.cnki.issn1003-207x.2007.06.020.
[13] Chen W, Chiu J, Chung H. Investor Trading Behavior, Market Liquidity, and the Role of Investor Sentiment[C]. 2009 FMA Annual Meeting. 2009.
[14] 李春红,彭光裕.投资者情绪、股票市场流动性和经济增长[J].技术经济,2011,30(05):105-108+123.
[15] 王丹枫,梁丹.从投资情绪角度看股票市场流动性——来自B股向境内居民开放的研究[J].数理统计与管理,2012,31(02):363-373.DOI:10.13860/j.cnki.sltj.2012.02.023.
[16] 刘晓星,张旭,顾笑贤,姚登宝.投资者行为如何影响股票市场流动性?——基于投资者情绪、信息认知和卖空约束的分析[J].管理科学学报,2016,19(10):87-100.
[17] 杨晓兰,沈翰彬,祝宇.本地偏好、投资者情绪与股票收益率:来自网络论坛的经验证据[J].金融研究,2016(12):143-158.
[18] 石广平,刘晓星,魏岳嵩.投资者情绪、市场流动性与股市泡沫——基于 TVP-SV-SVAR 模型的分析[J].金融经济学研究,2016,31(03):107-117.
[19] 张宗新,王海亮.投资者情绪、主观信念调整与市场波动[J].金融研究,2013,(04):142-155.
[20] 胡昌生,池阳春.投资者情绪、资产估值与股票市场波动[J].金融研究,2013,(10):181-193.
[21] 陆剑清,陈明珠.投资者情绪对股市波动影响的内在机制探析[J].山东社会科学,2015(S2):122-123.DOI:10.14112/j.cnki.37-1053/c.2015.s2.047.
[22] 巴曙松,朱虹.融资融券、投资者情绪与市场波动[J].国际金融研究,2016(08):82-96.DOI:10.16475/j.cnki.1006-1029.2016.08.008.
[23] 童元松.投资者情绪、股市流动性与波动性的关系研究[J].技术经济与管理研

究,2021(02):76-82.
[24] 何基报,鲁直.什么影响着投资者选择卖出或继续持有?[J].管理科学学报,2006(06):53-63.
[25] Pang B, Lee L. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. In: Scott D, ed. Proc. of the ACL 2004. Morristown: ACL, 2004. 271-278.
[26] 王靖一,黄益平.金融科技媒体情绪的刻画与对网贷市场的影响[J].经济学(季刊),2018,17(04):1623-1650.DOI:10.13821/j.cnki.ceq.2018.03.15.
[27] . Al-Nasseri A,Ali F M.What Does Investors’ Online Divergence of Opinion Tell Us About Stock Returns and Trading Volume?[J].Journal of Business Research,2018,86(1).
[28] . Pawar K,Jalem R S,Tiwari V.Stock Market Price Prediction Using LSTM RNN[M]Emerging Trends in Expert Applications and Security.Singapore:Springer,2019:493-503.
[29] DEVLIN J, CHANG M W, LEE K, et al. Bert: pre-training of deep bidirectional transformers for language understanding[C]//Proceedings of the 2019 Conference of the North American Chapter of the Association for ComputationalLinguistics:Human Language Technologies. Stroudsburg, PA: Association for Computational Linguistics, 2019: 4171-4186.
[30] Antweiler， W.， Frank， M.Z.Is All That Talk Just Noise? The Information Content of Internet Stock Message Boards[J].The Journal of Finance， 2004， 59（3）：1259-1294
[31] Nakajima, J. Time-varying parameter VAR model with stochastic volatility: An overview of methodology and empirical applications[J]. Monetary and Economic Studies, 2011(29): 107-142.
[32] 郑挺国,刘堂勇.股市波动溢出效应及其影响因素分析[J].经济学(季刊),2018,17(02):669-692.
[33] 杨文祺,李珂涵.投资者情绪和市场流动性如何影响股票市场稳定性?——基于股价崩盘视角的科创板市场TVP-VAR模型时变研究[J].制度经济学研究,2020(04):73-93.
[34] 任永平,李伟.经济政策不确定性、投资者情绪与股价同步性——基于TVP-VAR模型的时变参数[J].上海大学学报(自然科学版),2020,26(05):769-781.